

PAT-NO: JP409044639A  
DOCUMENT-IDENTIFIER: JP 09044639 A  
TITLE: METHOD AND DEVICE FOR CLASSIFYING VIDEO BLOCKS  
PUBN-DATE: February 14, 1997

INVENTOR-INFORMATION:

NAME

NIIKURA, YASUMASA  
HAMADA, HIROSHI  
AKUTSU, AKITO  
TANIGUCHI, YUKINOBU

ASSIGNEE-INFORMATION:

NAME

NIPPON TELEGR & TELEPH CORP <NTT>

COUNTRY

N/A

APPL-NO: JP07197416

APPL-DATE: August 2, 1995

INT-CL (IPC): G06T001/00, H04N005/262

ABSTRACT:

PROBLEM TO BE SOLVED: To automatically classify images into sets of similar shots without finely dividing the image for the unit of a shot and without damaging time order property.

SOLUTION: First of all, the feature amounts of respective shots are calculated and the shot is used as the block in an initial state (step 111). The degree of similarity between adjacent blocks is calculated (step 112), the adjacent block or step showing the maximum degree of similarity among the calculated degrees of similarity is merged with one block (step 113) and the feature amount of the merged block is calculated (step 114). The

**THIS PAGE BLANK (USPTO)**

steps 112-114  
are repeated until the classification is sufficiently performed, and  
the number  
of blocks is gradually decreased..

COPYRIGHT: (C)1997,JPO

**THIS PAGE BLANK (USPTO)**

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-44639

(43)公開日 平成9年(1997)2月14日

(51)Int.Cl. <sup>8</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 T 1/00			G 0 6 F 15/62	P
H 0 4 N 5/262			H 0 4 N 5/262	

審査請求 未請求 請求項の数7 O L (全12頁)

(21)出願番号 特願平7-197416

(22)出願日 平成7年(1995)8月2日

(71)出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72)発明者 新倉 康巨

東京都千代田区内幸町一丁目1番6号 日

本電信電話株式会社内

(72)発明者 浜田 洋

東京都千代田区内幸町一丁目1番6号 日

本電信電話株式会社内

(72)発明者 阿久津 明人

東京都千代田区内幸町一丁目1番6号 日

本電信電話株式会社内

(74)代理人 弁理士 若林 忠

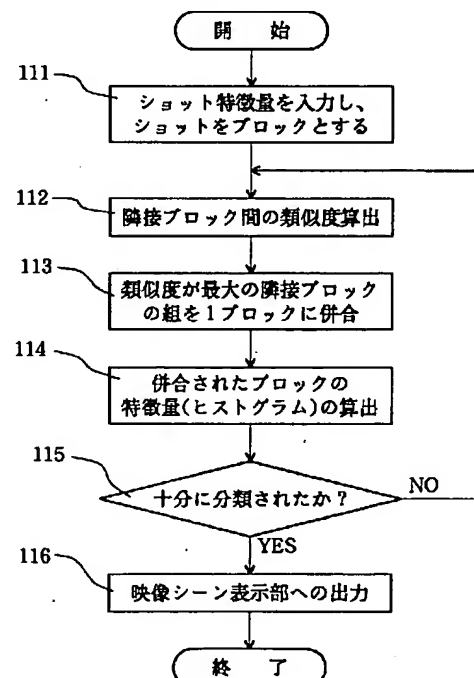
最終頁に続く

(54)【発明の名称】 映像ブロック分類方法及び装置

(57)【要約】

【課題】 映像をショット単位に細分することなく、かつ時間的な順序性を損なうことなく、類似したショットの集合に映像を自動的に分類する。

【解決手段】 まず、各ショットの特徴量を算出し、初期状態のブロックとしてショットを用いる(ステップ111)。隣接するブロック間の類似度を算出し(ステップ112)、算出された類似度の中で最大の類似度を示した隣接するブロックないしステップを1つのブロックに併合し(ステップ113)、併合されたブロックの特徴量を算出する(ステップ114)。十分に分類が行われるまで、ステップ112～114を繰り返し、ブロックの数を漸減させる。



1

## 【特許請求の範囲】

【請求項1】 画像データ列で構成された映像を複数の映像ブロックに分類する映像ブロック分類方法であって、

予め多数のブロックに分類された映像を入力する映像入力工程と、

各ブロックの画像データ列から当該ブロックの特徴量を算出する特徴量算出工程と、

特徴量に基づいて隣接するブロック間の類似度を算出する類似度算出工程と、

算出された類似度の中で最大の類似度を示した隣接するブロックを1つのブロックに併合するブロック併合工程と、を有し、

特徴量算出工程、類似度算出工程及びブロック併合工程を反復して実行することにより、分類された複数の映像ブロックを得る映像ブロック分類方法。

【請求項2】 映像入力工程が、フレーム単位で入力した映像をブロックの1種であるショットに分類する工程を含む、請求項1に記載の映像ブロック分類方法。

【請求項3】 各ブロックの特徴量が、当該ブロックに含まれるフレームから色、色相、彩度、明度のいずれかあるいはこれらの組み合わせからなる情報を抽出して得たヒストグラムであり、当該ブロックに含まれるフレーム数が2以上の場合には、当該ブロックに含まれるフレームの全ヒストグラムについてのヒストグラム論理積演算を行って得たヒストグラムとして表わされる、請求項1または2に記載の映像ブロック分類方法。

【請求項4】 類似度が、隣接するブロックのヒストグラムに対するヒストグラム累積論理積に基づいて定められる、請求項3に記載の映像ブロック分類方法。

【請求項5】 ブロック併合工程における隣接するブロックの併合に際し、当該隣接するブロックのヒストグラムに対してヒストグラム論理積演算を行い、演算の結果得られたヒストグラムを併合後のブロックの特徴量とする請求項3または4に記載の映像ブロック分類方法。

【請求項6】 画像データ列で構成された映像を複数の映像ブロックに分類する映像ブロック分類装置であって、

入力された画像データ列を保存する画像データ列メモリと、

画像データ列メモリからフレームのデータを読み出し、フレームごとのフレーム特徴量を算出する画像情報変換部と、

画像データ列メモリからフレームを読み出し、ショット単位に分類するショット単位分類部と、

フレーム特徴量に基づいて各ショットの特徴量を算出するショット特徴量算出部と、

1または複数のショットで構成されるブロックに対し、ブロックに対する特徴量を利用して隣接するブロック間の類似度を算出し、算出された類似度の中で最大の類似

2

度を示した隣接するブロックを1つのブロックに併合することによって複数のショットで構成されるブロックを生成し、類似度の算出とブロックの併合とを繰返し実行する類似度評価及び画像分類処理部と、を有する映像ブロック分類装置。

【請求項7】 フレーム特徴量及びブロックの特徴量がそれぞれヒストグラムで表わされ、ショットごとにフレーム特徴量のヒストグラム論理積を算出することで各ショットの特徴量が算出され、隣接するブロックのヒストグラムのヒストグラム累積論理積に基づいて類似度が算出され、類似度評価及び画像分類処理部が、隣接するブロックの併合に際して当該隣接するブロックのヒストグラムに対してヒストグラム論理積演算を行い演算の結果得られたヒストグラムを併合後のブロックの特徴量とする、請求項6に記載の映像ブロック分類装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、映像データの処理に関し、特に、映像を構成する画像データ列を分類してブロックやシーン単位に構造化する方法及び装置に関する。

## 【0002】

【従来の技術】映像データは一般にデータ量が膨大であるが、その内容を知るためには、映像を時間順に全て見ていくしかなかった。時間的に連続する画像データの集合が映像であると考え、映像を構成する各画像データがある尺度に基づいてそれぞれ分類されていれば、映像の概略を把握したり、いわゆる飛ばし見をしたりするのに有用であり、短時間で映像の内容を理解するのに非常に役立つのではないかと期待される。以下、映像の分類とは、映像に内在する階層的な構造要素を検出し、同一の分類に属する構造要素を束ねることによって、時間軸に関して映像を複数のブロックに分割することを意味する。同一の分類に属する構造要素を束ねることに着目すれば、映像の構造化ともいうことができる。

【0003】一般に映像は、「フレーム」、「テイク」、「ショット」、「シーン」、「ストーリー」という単位で階層的に分類される。この階層的分類には、ハードウェアでの物理的なレベル（下位レベル）から、人間の創造的作業による意味的なレベル（上位レベル）までの全てを含んでいる。

【0004】「フレーム」は、映像を撮影するときのフィルム1コマ1コマに対応する物理的な単位であり、「テイク」は、フレームの集合であって同一カメラで撮影された時間的に連続な映像区間を示す単位であり、これには、撮影時のカメラのオン（ON）／オフ（OFF）が反映される。

【0005】「ショット」は、テイクと同様に同一カメラで撮影された映像区間を示す単位であるが、映像の編集時にテイクの中から選び出された映像区間をいい、テ

イクよりもより人間の意図が反映された意味的な単位である。さらに、編集作業によりショットが組み合わされると、映像の「ストーリー」となる。すなわち、テイクは映像素材の単位であり、ショットは映像作品の構成単位である。

【0006】一方、「シーン」とは、映像作品において、意味的に同一の場面とみなせるショットの集合を指し、映像作品を幾つかのブロックにわけた際の単位となる。シーンの分類には人間の創造的な作業を多く必要とし、さまざまな解釈によっていろいろに分類される。例として、登場人物が一致していれば、同一シーンとする解釈もあるだろうし、時間的にどれだけ離れていても、映像を撮影した舞台や場所が一致していれば、同一シーンとする場合もある。しかしここでは、作品の内容把握を助けるために行うためにシーンの分類を行うという視点にたち、シーンに以下の条件を設ける。

条件1：シーンは、似たような画像特性をもつショットの集合である（以下、特徴類似条件という）。

条件2：シーンは、時間的に隣接し連続したショットの集合である（以下、時間的連続条件という）。

【0007】以上の条件を満たすシーンの最も端的な例は、同一被写体を、複数の異なるカメラで異なる視点から撮影した映像から、それらの切り替えの連続によって作成された映像である。すなわち、類似し、かつ、連続な複数のショットによって一つのシーンを構成している例である。一方で、例えばニュース番組映像によく見られるスタジオと現場の中継の映像が交互に存在するような映像は、それぞれ、スタジオの場面同士は類似し、現場の中継同士も類似しているが、スタジオと現場中継映像の間の類似性は低いので、同一のシーンであるとはしない。類似した映像が連続的に構成されているわけではないので、このような場合は、スタジオシーンA、現場シーンA、スタジオシーンB、現場シーンB、スタジオシーンC、現場シーンC、…というように、それぞれ全く別個のシーンが連続に続いているとみなす。

【0008】理想的な分類を行うにはストーリーなどまでを考慮しなければならないが、現状ではこの作業は人手によってしか行うことができず、作業量が膨大となって、特別の場合を除いて非現実的なものとなる。したがって、画像データの分類に関し、なんらかの自動化が求められている。

【0009】上述した映像の分類単位に基づきユーザが利用しやすいように画像データを分類して映像を構造化することを目的とした従来技術がいくつか存在する。

【0010】例えば、連続する画像フレームのフレーム間での対応する位置(x, y)の輝度の差分の総和から、連続するフレームにおける変化率を計算して映像のカット点の切り替わりを検出する「映像カット点検出方法」等があげられる。これはショット単位の分類技術としてとらえられる。

【0011】シーン単位の分類を目的とした技術としては、(1)同一のシーンは似た色情報からなるという立場から、画像データ列の色情報を特徴空間へ変換し、特徴空間上でクラスタリングし、映像の分類を行う「映像特徴処理方法」(特開平6-251147号)や、(2)一般に映像作品においては画像とともに映像データを構成する要素である音情報がショット単位でなくシーン単位にかつ意味的な作業によって付加されていることを利用して、この音情報によってシーンを分類する「音情報を用いたビデオ・ブラウジング・インタフェース」(テレビジョン学会技術報告, Vol.19, No.7, 1995/12)や、(3)ショットごとの代表画面を求め、代表画像間での輝度のモーメント不変量と色情報とを基に類似度を算出し、類似性の高い代表画面及びショットを表示し、低いものを表示しないインタフェースを作成し、結果的にユーザに対して類似した代表画面を表示することによって映像をブロックに分ける「Content-based Browsing of Video Sequences」(ACM, Multimedia 94, P.97-)等が存在する。

20 【0012】

【発明が解決しようする課題】映像の内容の把握を目的として、映像を分類する場合には、意味を反映するブロックすなわちシーン単位に分類されることが望ましい。

【0013】上述した従来の技術のうち、「映像カット点検出方法」は、映像におけるカット点の検出を目的とし、ショット単位での分類を可能にしている。しかし、例えば2時間の映像作品は一般に数千にも及ぶショットから構成されており、ショット単位では、映像の内容把握のためには細分化されすぎてしまうという問題点が生じる。したがって、ショット単位ではなくシーン単位に分類する技術が必要となる。

30

【0014】一方、「映像カット点検出方法」を除いた他の従来技術は、ショット単位の分類では細分化されすぎる点を解決すべく、より意味的な単位に分類を行うことを目的としたものである。このうち、「映像特徴処理方法」では、同一のシーンは類似した色情報によって構成されているという仮定に基づき、色情報に基づく特徴空間を使用して、類似した色の組み合わせをもつ画像の分類を行っている。この「映像特徴処理方法」は、上述の特徴類似条件(条件1)を満たしているが、時間的に連続であるという時間的連続条件(条件2)を考慮していない。したがって、この方法によれば、類似する画像群を抽出して、時間的に不連続でかつ類似した画像を1つのブロックとして検出してしまうことがある。すなわち、上述のニュース番組映像の例を用いて説明すれば、時間的に不連続なスタジオでの映像をそれぞれ個別のシーンとして分類することなく、一つの類似したブロックとみなしてしまう。結局、この方法では、時間的に連続なシーンを安定に抽出することができない。

50 【0015】「音表現を用いたビデオ・ブラウジング・

5

インタフェース」は、BGM等の音情報を利用することによって、細分化されたショットを一連の時間的に連続な同一シーンに併合することができ、上述の時間的連続条件を満たす分類を行うことができる。しかしながら、シーンをまたがって同じBGMが連続している場合やシーンの途中から音情報が挿入された場合、さらには音情報が全く存在しないような映像作品等のように、シーンと音情報が必ずしも対応していない場合には、安定して映像をシーンに分類することができないという問題点がある。さらに、特徴類似条件を満たさないという問題点がある。

【0016】「Content-based Browsing of Video Sequences」では、予めショット単位に分類を行い、ショットごとに代表画像を選び出し、これらの代表画像同士のモーメント不変量の比較と色情報の比較との両方を利用して類似度を判定している。形状と色による類似度を用いることによって、代表画面を選択した際に類似する代表画面のみを表示することにより、注視している画像に類似し関連する情報の簡単な表示、検索を実現している。したがって、「映像特徴処理方法」と同様に特徴類似条件は満たすが、時間的連続条件は満たさない。

【0017】映像の内容把握を的確に行えるような分類を実行するためには、上述の特徴類似条件と時間的連続条件の両方を満たすシーン分類を実行しなければならないが、以上述べた従来の技術にはこれら2条件を同時に満足するものはない。

【0018】本発明の目的は、映像をショット単位に細分することなく、かつ時間的な順序性を損なうことなく、類似したショットの集合に映像を分類すること、すなわち、特徴類似条件と時間的連続条件とを同時に満たし、映像の分類を安定して行うことができる映像ブロック分類方法及び装置を提供することにある。

【0019】

【課題を解決するための手段】本発明の映像ブロック分類方法及び装置は、特徴類似条件と時間的連続条件とを満たしつつ映像を分類するために、以下の構成を有する。

【0020】すなわち本発明の映像ブロック分類方法は、画像データ列で構成された映像を複数の映像ブロックに分類する映像ブロック分類方法であって、予め多数のブロックに分類された映像を入力する映像入力工程と、各ブロックの画像データ列から当該ブロックの特徴量を算出する特徴量算出工程と、特徴量に基づいて隣接するブロック間の類似度を算出する類似度算出工程と、算出された類似度の中で最大の類似度を示した隣接するブロックを1つのブロックに併合するブロック併合工程と、を有し、特徴量算出工程、類似度算出工程及びブロック併合工程を反復して実行することにより、分類された複数の映像ブロックを得る。

【0021】本発明の映像ブロック分類方法において、

6

映像入力工程に、フレーム単位で入力した映像をブロックの1種であるショットに分類する工程を含ませてもよい。また、各ブロックの特徴量としては、当該ブロックに含まれるフレームから色、色相、彩度、明度のいずれかあるいはこれらの組み合わせからなる情報を抽出して得たヒストグラムを用いることが好ましく、この場合、ブロックに含まれるフレーム数が2以上の場合には、当該ブロックに含まれる映像フレームの全ヒストグラムについてのヒストグラム論理積演算を行って得たヒストグラムを特徴量とすることが好ましい。また、類似度は、隣接するブロックのヒストグラムに対するヒストグラム累積論理積に基づいて定めることが好ましい。さらに、ブロック併合工程における隣接するブロックの併合に際し、当該隣接するブロックのヒストグラムに対してヒストグラム論理積演算を行い、演算の結果得られたヒストグラムを併合後のブロックの特徴量とすることが好ましい。

【0022】本発明の映像ブロック分類装置は、画像データ列で構成された映像を複数の映像ブロックに分類する映像ブロック分類装置であって、入力された画像データ列を保存する画像データ列メモリと、画像データ列メモリからフレームのデータを読み出し、フレームごとのフレーム特徴量を算出する画像情報変換部と、画像データ列メモリからフレームを読み出し、ショット単位に分類するショット単位分類部と、フレーム特徴量に基づいて各ショットの特徴量を算出するショット特徴量算出部と、1または複数のショットで構成されるブロックに対し、ブロックに対する特徴量を利用して隣接するブロック間の類似度を算出し、算出された類似度の中で最大の類似度を示した隣接するブロックを1つのブロックに併合することによって複数のショットで構成されるブロックを生成し、類似度の算出とブロックの併合とを繰返し実行する類似度評価及び画像分類処理部と、を有する。

【0023】本発明の映像ブロック分類装置では、フレーム特徴量及びブロックの特徴量をヒストグラムで表わし、ショットごとにフレーム特徴量のヒストグラム論理積を算出することで各ショットの特徴量を算出し、隣接するブロックのヒストグラムのヒストグラム累積論理積に基づいて類似度を算出し、類似度評価及び画像分類処理部が、隣接するブロックの併合に際して当該隣接するブロックのヒストグラムに対してヒストグラム論理積演算を行い演算の結果得られたヒストグラムを併合後のブロックの特徴量とすることが、好ましい。

【0024】結局、本発明の映像ブロック分類方法及び装置では、類似する特徴量、典型的には類似する色情報をもった隣接する画像ブロックを、類似度の高いものから併合するという処理を反復している。このため、特徴類似条件と時間的連続条件というシーンの条件を満たし、類似した画像の集合でかつ時間的に連続な画像ブロックの抽出が可能である。また、分類された映像ブロッ



クをさまざまな形でユーザに表示・提供し、ユーザからの入力を受け取るインタフェースを充実することによって、ユーザによる映像内容の把握を助けることが可能になる。

【0025】

【発明の実施の形態】以下、図面を用いて本発明の実施の形態を説明することにより、本発明をさらに詳しく説明する。

【0026】この実施の形態では、ショット単位などで多数のブロックに分割された映像について、各ブロックごとにそのブロックの映像の特徴を表わすヒストグラムを求め、ヒストグラムに基づき時間的に隣接する2ブロック間の類似度を算出し、類似度が最大となった隣接する2ブロックを1つのブロックに併合する。この処理を繰り返すことによって、ブロック数が漸減する。十分に分類が行われたかの判定を行い、その結果、十分に分類が行われたときにもとの映像がシーンに分類される。

【0027】ヒストグラムは、例えば、画像データから、色、色相、彩度、明度などの特徴量を抽出してこの特徴量を成分順に配置したものである。画像データにおける色や明度などの空間分布を直交変換して得た周波数分布をそのままヒストグラムとしてもよい。そしてブロックの映像を表わすヒストグラムは、そのブロックに属する各フレームのヒストグラムを得た上で、全フレームのヒストグラムについてのヒストグラム論理積を計算することによって得られる。ヒストグラム論理積の演算は、後述する説明から明らかになるように、ヒストグラムにおける各成分ごとに、論理積の対象となるヒストグラムにおけるその成分の値の最小値を求め、各成分ごとの最小値を並べてヒストグラムを生成することによって実行される。したがって、ブロックの併合に際しては、併合対象のブロックのヒストグラムを対象としてヒストグラム論理積の演算を行ってヒストグラムを求めることにより、併合後のブロックのヒストグラムを得ることができる。また、隣接ブロック間の類似度としては、隣接ブロックのヒストグラムのヒストグラム論理積の演算を行って得たヒストグラムの面積、すなわち後述するヒストグラム累積論理積を使用する。

【0028】図1は、本発明の実施の一形態の映像ブロック分類装置の構成を示すブロック図である。この映像ブロック分類装置は、入力画像データ列10をシーンに分類し、シーンに分類された映像16として映像シーン表示部17に表示するものであり、ユーザインタフェース部18を介して入力するユーザからの要求によって、分類の度合（最終的に分類されるシーンの数など）を調節できるようになっている。

【0029】入力画像データ列10における画像のサンプリングレート、画像のデータフォーマット、画像サイズは任意である。例えば、入力画像データ列10は、NTSC標準映像信号を30フレーム/秒でサンプリングした

ものであってもよいし、それよりも粗いサンプリングレートでサンプリングしたものであってもよい。また、入力画像データ列10は、NTSCのようなアナログ信号であってもデジタル信号であってもよく、ビデオカメラなどから直接入力されるデータであってもハードディスクやCD-ROM等の蓄積装置に保存されている画像ファイルであってもよい。図1に示した例では、入力画像データ列10は $t+1$ 枚のフレーム $I_0, I_1, \dots, I_t$ で構成されるNTSC映像信号である。

【0030】入力画像データ列10をフレーム単位で格納する画像データ列メモリ11が設けられている。画像データ列メモリ11は、単純に入力画像データ列を格納しておくだけでなく、ある程度まで加工されたデータを保存していても構わないし、ショットや撮影者の名前、撮影時の場所等の付加情報や、本実施の形態での以下に述べる処理の結果得られる情報などを、同時に格納してもよい。ここでは、画像データ列メモリ11には入力画像データ列の原信号を保存するものとする。

【0031】画像データ列メモリ11から読み出された入力画像データ列が入力するショット単位分類部12及び画像情報変換部13が設けられている。ショット単位分類部12は、ショット単位分類処理を実行して、各フレーム $I_0, I_1, \dots, I_t$ のデータに基づき入力画像データを $n+1$ 個（ただし $n < t$ ）のショット $S_0, I_1, \dots, S_n$ に分類するものである。ショット単位分類処理は、映像信号に予め付加されるショット情報を利用して構わないし、既存のカット点検出技術を利用してショットに分類しても構わない。また、人間が予めショットに分類しておいてもよい。本実施の形態では、既存のカット点検出技術によってショット単位に分類する。なお、連続する1あるいは複数のフレームで構成するものを一般的にブロックというから、ショットもそしてシーンもそれぞれブロックの1種である。

【0032】画像情報変換部13は、各フレームの画像情報を、色、色相、彩度、明度等の情報に変換し、フレーム特徴量 $H_0, H_1, \dots, H_t$ を生成するものである。フレーム特徴量は、色、色相、彩度、明度等の情報に基づくヒストグラムとして表わされている。色、色相、彩度、明度等の情報以外の他の情報に変換しても構わない。ここでは、画像フレーム $I_0, I_1, \dots, I_t$ の全てをそれぞれRGB情報に変換し、RGBヒストグラムとしてフレーム特徴量 $H_0, H_1, \dots, H_t$ を出力している。

【0033】そして、ショット特徴量算出部14が設けられ、ショット特徴量算出部14は、ショット単位分類部12で分類されたショット $S_0, S_1, \dots, S_n$ の情報と画像情報変換部13で得られたフレーム特徴量 $H_0, H_1, \dots, H_t$ に基づいて、ショット $S_0, S_1, \dots, S_n$ ごとにそのショット内の全フレームのヒストグラム論理積を演算して各ショットごとの特徴量すなわちショット特徴量 $S_{H_0}, S_{H_1}, \dots, S_{H_n}$ を算出する。算出されたショット特

微量 $SH_0, SH_1, \dots, SH_n$ は、類似度評価及び画像分類処理部15に入力する。類似度評価及び画像分類処理部15は、それぞれヒストグラムとして表わされるブロックの特徵量（ブロックの特徵量には、ショット特徵量 $SH_0, SH_1, \dots, SH_n$ や、ショットを併合したブロックの特徵量が含まれる）を基に、ヒストグラム累積論理積から、隣接するブロック（ここでのブロックにはショットも含まれる）の類似度を計算して評価し、類似度が最大の隣接する組み合わせを併合して新たなブロックを構成するものである。類似度評価及び画像分類処理部15は、ブロックないしショットの併合によって新たなブロックを生成する際、併合されたブロックないしショットの各特徵量のヒストグラム論理積に基づき、併合後のブロックの特徵量（ヒストグラム）を算出する。実際には、ユーザインタフェース18部からの指示に応じて、類似度評価及び画像分類処理部15はこの評価、併合、特徵量算出という処理を反復して実行し、シーンに分類された映像16を映像シーン表示部17に出力する。なお、ユーザは、映像シーン表示部17に表示されるシーンに分類された映像16に対し、時間方向により詳細にシーンを表示したいなどの要求をユーザインタフェース部18から入力することによって、その要求を映像に反映させることが可能である。

【0034】ショット単位分類部12、画像情報変換部13、ショット特徵量算出部14及び類似度評価及び画像分類処理部15は、演算能力をもつCPUを利用したソフトウェアによって処理を実現してもよいし、複数のCPUとソフトウェアの組み合わせによって実現してもよいし、一部を専用のハードウェアによって実現してもよいし、全部を専用のハードウェアを用いて実現してもよい。ここでは、演算能力をもつCPUを利用したソフトウェアによって処理を実現している。

【0035】次に、ショット特徵量の算出について、図2を用いて説明する。画像処理変換部13では、各フレームのデータから色データ（RGBデータ）の入力が行われ（ステップ101）、フレームごとにフレーム特徵量のヒストグラム $H_0, H_1, \dots, H_t$ が生成し（ステップ102）、ショット特徵量算出部14に入力する。一方、それぞれのショット $S_0, S_1, \dots, S_n$ にはどのフレームが属するかの情報も、ショット単位分類部12からショット特徵量算出部14に入力している。そこでショット特徵量算出部14は、ショット $S_0, S_1, \dots, S_n$ ごとにそのショットに含まれるフレーム特徵量（ヒストグラム）のヒストグラム論理積を算出し（ステップ103）、ヒストグラム論理積の演算で得られたヒストグラムをショットごとにそのショット特徵量 $SH_0, SH_1, \dots, SH_n$ として出力する（ステップ104）。

【0036】ここで、ヒストグラム論理積の演算の詳細について具体的に説明する。上述したようにヒストグラム論理積の演算は、ショット特徵量の算出のみならず、

類似度の算出や併合されたブロックの特徵量の算出などに使用されるものである。

【0037】まず、 $H_1, H_2, H_3, \dots$ を各ヒストグラムとし、 $H_1(j), H_2(j), H_3(j), \dots$ を成分jにおける各ヒストグラム $H_1, H_2, H_3, \dots$ の値とする。また、ヒストグラム論理積の計算の結果得られるヒストグラムを $H_{new}$ とし、この新たなヒストグラム $H_{new}$ での成分jに対する値を $H_{new}(j)$ とする。

10 【0038】各ヒストグラムの成分jが0からmまで存在し、かつ、ヒストグラム論理積を算出するためのヒストグラムが1からkまで存在するとき、ヒストグラム論理積の演算は、

【0039】

【数1】 $H_{new}(j) = \min \{H_1(j), H_2(j), \dots, H_k(j)\}$

（ただし $0 \leq j \leq m$ ）で表わされる。

【0040】図3は、図示(a),(b)に示されるように隣接する2つのブロック（ブロックA,B）に対応するヒストグラムが与えられたときに、どのようにこれら2つのブロックA,B間のヒストグラム論理積が計算されるのかを図解したものである。図示(c)はブロックA,B間のヒストグラム論理積を示している。すなわち、特徴成分ごとに、各ブロックでの値のうち低い方の値がヒストグラム論理積演算の結果のヒストグラムに採用されている。例えば、成分Kについては、図示実線矢印で示すように、ブロックBの方が値が小さいので、ヒストグラム論理積における成分Kの値はブロックBの値と等しくなる。同様に、成分Lについては、ブロックAの方の値が採用されている。

30 【0041】次に、類似度評価及び画像分類処理部15での処理について、図4を用いて説明する。類似度評価及び画像分類処理部15の処理は、簡単に言えば、ショットを含むブロックを併合する処理であり、その併合の過程において隣接する画像ブロックの特徵量同士を比較して類似度を算出し、その類似度に基づいて映像を小さなブロックから大きなブロックへと併合し、最終的には映像を例えばシーンに対応するブロックに分類する処理である。

40 【0042】類似度評価及び画像分類処理部15での併合処理の初期段階では、各ブロックはそれぞれ1つのショットで構成されているはずである。そこでまず、ショット特徵量算出部14から各ショット $S_0, S_1, \dots, S_n$ のショット特徵量 $SH_0, SH_1, \dots, SH_n$ を入力してこれらをここでの処理対象のブロックとする（ステップ111）。次に、ヒストグラム累積論理積に基づいて、隣接する2つのブロック間の類似度を算出する（ステップ112）。隣接する2つのブロック（ショットも含む）の特徵量を示すヒストグラムをそれぞれ $H_1, H_2$ とし、ヒストグラム $H_1, H_2$ での成分jの値をそれぞれ

11

れ $H_1(j)$ 、 $H_2(j)$ とする。また、ヒストグラム累積論理積（すなわち類似度）を $V$ とする。ヒストグラムの成分 $j$ が0から $m$ までであるとする、ヒストグラム累積論理積 $V$ は、

【0043】

【数2】

$$V = \sum_{j=0}^m \min(H_1(j), H_2(j))$$

（ただし $0 \leq j \leq m$ ）で表わされる。

【0044】図5は、図示(a)、(b)に示されるように隣接する2つのブロック（ブロックA、B）に対応するヒストグラムが与えられたときに、どのようにこれら2つのブロックA、B間のヒストグラム累積論理積が計算されるのかを図解したものである。ヒストグラム論理積の演算と同様に、成分ごとにヒストグラムでの値を比較し、低い方の値をその成分における値とし、さらにその累積を求めている。すなわち、ヒストグラム累積論理積 $V$ は、両者間のヒストグラム論理積を示す図示(c)における斜線部の面積である。各成分の累積和によって面積を得ているため、ここではヒストグラム累積論理積と呼んでいるのである。

【0045】隣接する2ブロック間のヒストグラム論理積の面積の大きい場合、すなわちヒストグラム累積論理積が大きい場合には、隣接する2つのブロック間で同じ成分に対する値がそれぞれ大きいことが反映されており、特徴量における各成分レベルで画像が類似しているといえることができる。したがって、ヒストグラム累積論理積によって、隣接する画像ブロックの類似度の評価を行うことができるのである。

【0046】上述したようにヒストグラム累積論理積に基づく類似度は、値が大きいほど隣接する2つのブロック間での隣接の度合いが高いことを示している。そこで、隣接するブロック間で求めた類似度の中で最大の類似度を探索し、この最大の類似度を示した隣接する2ブロックを1つのブロックに併合する（ステップ113）。そして、併合によって新たに生成した画像ブロックの特徴量を算出する（ステップ114）。ここでは併合前の2ブロックのヒストグラム論理積を併合後のブロックの特徴量とする。その後、映像の分類が十分に進行したかどうかを判定し（ステップ115）、十分に分類された場合にはシーンに分類された映像を16を映像シーン表示部17に出力して処理を終了し、十分でない場合には、ステップ112に戻り、併合されたブロックを対象として隣接ブロック間の類似度の算出を実行する。なお、併合されていないブロック間の類似度は変化しないから、2回目以降にステップ112を実行する場合には、直前に併合されて生成したブロックとこのブロックに隣接するブロックとの類似度のみを算出すればよい。

【0047】分類が十分にいくかどうかの判定は、ブロック間での類似度を参考にして行うことができる。この

12

実施の形態では、ブロックの特徴量であるヒストグラムの算出を、そのブロックを構成するより小さなブロックの特徴量であるヒストグラム同士の論理積によって算出している。そのため、併合が繰り返された後に得られるブロックは、多数のフレームないしショットによって構成されるが、一方でその特徴量であるヒストグラムの値は小さいものとなる。したがって、巨大なブロック同士の類似度を示すヒストグラム同士の論理積の値は、双方のヒストグラムの値がきわめて小さなものであるため、非常に小さい値でしかない。このことを用いて、ヒストグラム累積論理積の値が0のときを全く類似していない状態、すなわち映像が十分に分類された状態とし、ヒストグラム累積論理積の値が0でないものが存在する場合は、分類が十分に行われていない状態であって分類処理を継続すべき場合であるといえることができる。また、この実施の形態は、単純な反復作業のみによって実現されているため、シーンの分類レベルをさまざまな形で調節することが可能である。例えば、ユーザインタフェース部18を介してユーザから入力する要求にしたがって分類レベルを変化させ、最終的に分類されるブロックの個数を変化させることが可能であり、ユーザからの入力情報を基に分類が十分に行われたかどうかを判断することもできる。

【0048】以下、図6を用いて、この実施の形態においてショットが併合されてより大きなブロックに分類されていく過程をより詳細に説明する。

【0049】(a)は、ショット単位分類部12によって分類された5つのショットA～Eを示している。これらのショットA～Eから、ユーザにとって好適な分類単位であるシーン単位の分類を進める。ショット特徴量算出部14によって算出されたショットA～Eごとのショット特徴量（ヒストグラム）が(b)に示されている。ショットに含まれる各フレームのヒストグラム（特徴量）から、ヒストグラム論理積を計算することによって、ショット特徴量が得られている。

【0050】(c)は、隣接するショット間のヒストグラム論理積を示すことによって、隣接するショット間の類似度を算出する過程を示している。図において記号

【0051】

【外1】

⊗

は、ヒストグラム論理積の演算を示している。実際の類似度の評価では、ヒストグラム論理積の面積、すなわちヒストグラム累積論理積を用いている。類似度の評価の結果、ショットDとショットEとの組み合わせが最大の類似度を示したので、(d)に示すように、ショットDとショットEを併合してDEという併合ブロックを得る。

【0052】この実施の形態では、十分に映像がシーンに分類されるまで処理を行うので、併合ブロックDEを含むショット群に対してさらに処理を実行する。この時

## 13

点での各ショット及び併合ブロックに対する特徴量が(e)に示されている。そして、(f)に示すように、隣接するショットないしブロックの類似度をヒストグラム累積論理積によって計算する。実際には

【0053】

【数3】

$$A \otimes B, B \otimes C$$

は前回の計算値をそのまま使用し、

【0054】

【数4】

$$C \otimes D E$$

のみを新たに計算する。その結果、ショットCとブロックDEとの類似度が最大となったので、(g)に示すように併合ブロックCDEを生成する。以下同様の処理を繰返し、例えば完全にシーンに分類されるまで、特徴量算出、類似度算出、併合ブロック決定という処理を繰返し行っていけばよい。

【0055】以上、本発明の実施の形態について説明したが、ここでは、類似度の評価において1回の評価で1回の併合しか行われぬ。したがって、例えば、ユーザインタフェース部18を介して、より詳細に映像ブロックを見たい、については、「映像をn個数に分類したものが欲しい」といった要求があった場合には、この要求に対して瞬時に対応することが可能である。すなわち、図6の例においては、(a)の5つのショットA～Eからなる映像を4個に分類して見たいという要求があった場合、シーンを求める際に4個のブロックの状態になっている図示(d)の状態、すなわちA、B、CDEのように分類された映像ブロックを提示すれば良い。

【0056】なお、上述の実施の形態では、ヒストグラム累積論理積に基づいて類似度を算出し、最大の類似度の隣接ブロックを併合していくという例を説明したが、類似度の評価と併合ルールには他にさまざまな方法が考えられる。また、初期状態のブロックがショットである場合を説明したが、1ショットないし1テイク内の映像を分類するような場合には、初期状態でのブロックをフレームとして、上述の処理を行うことも可能である。

【0057】また本発明の応用として、データベースのブラウジングインタフェースや、映像コンテキストの制作など様々な映像処理と、ユーザインタフェースへの応用なども挙げられる。

【0058】

## 14

【発明の効果】以上説明ように本発明は、ショットやブロック単位に分類された映像を、隣接するショットあるいは画像ブロック間で類似度を算出し、評価し、併合を行うという処理を繰返し行うことにより、類似しているかつ時間的に連続な画像ブロックを集中して集めることが可能となり、特徴類似条件と時間的連続条件を満たす映像分類が実現できるという効果がある。したがって、複数の細分化されたショット単位に分類された映像をシーン単位に分類することが可能になる。

10 【0059】さらに、本発明は単純な反復処理によって映像の分類を実現しているため、分類のレベルを多様に調節することが可能であり、シームレスにシーンに分類された映像を提供できる。

【図面の簡単な説明】

【図1】本発明の実施の一形態の映像ブロック分類装置の構成を示すブロック図である。

【図2】ショット特徴量の算出の処理を説明するフローチャートである。

20 【図3】ヒストグラム累積論理積の算出方法を示す模式図である。

【図4】類似度評価及び画像分類処理部での処理を説明するフローチャートである。

【図5】ヒストグラム累積論理積の算出方法を示す模式図である。

【図6】ショットがブロックとしてまとめられていく過程を示す模式図である。

【符号の説明】

10 入力画像データ列

11 画像データ列メモリ

30 12 ショット単位分類部

13 画像情報変換部

14 ショット特徴量算出部

15 類似度評価及び画像分類処理部

16 シーンに分類された映像

17 映像シーン表示部

18 ユーザインタフェース部

101～104, 111～116 ステップ

A～E ショット

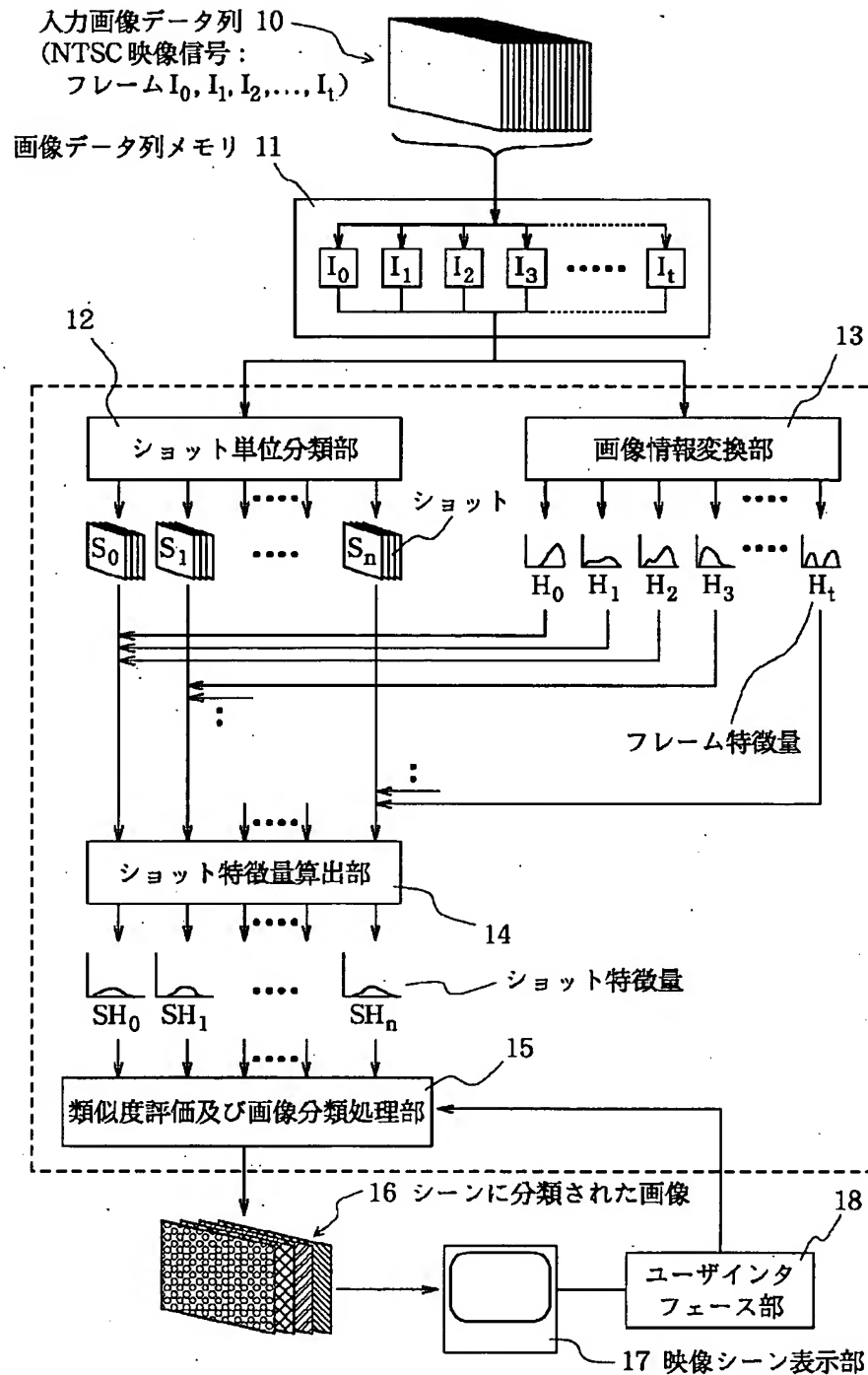
H<sub>0</sub>, H<sub>1</sub>, ..., H<sub>t</sub> フレーム特徴量

40 I<sub>0</sub>, I<sub>1</sub>, ..., I<sub>t</sub> フレーム

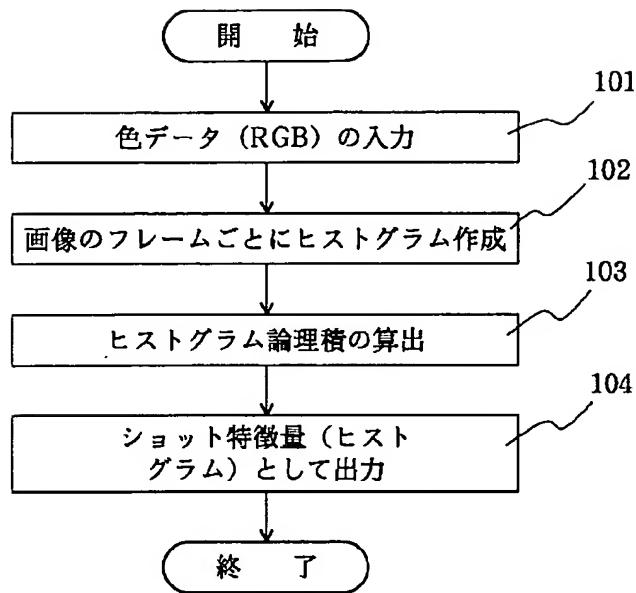
S<sub>0</sub>, S<sub>1</sub>, ..., S<sub>n</sub> ショット

SH<sub>0</sub>, SH<sub>1</sub>, ..., SH<sub>n</sub> ショット特徴量

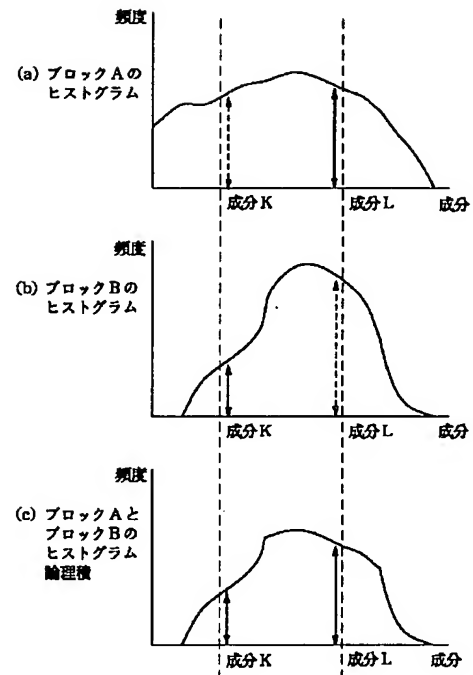
【図1】



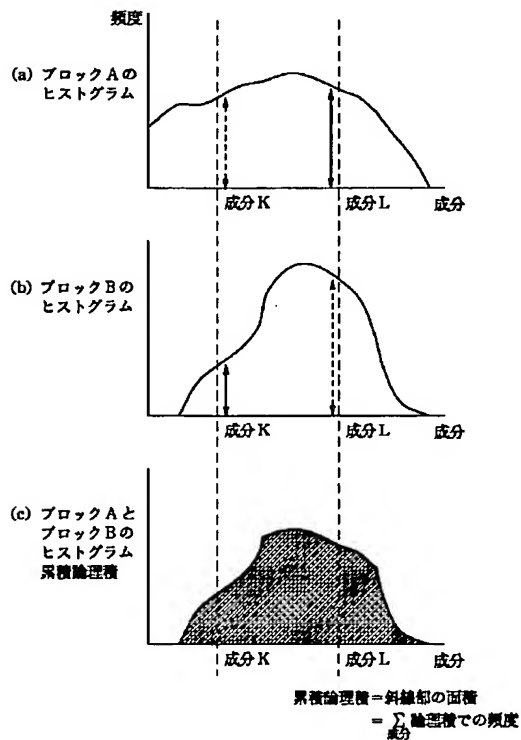
【図2】



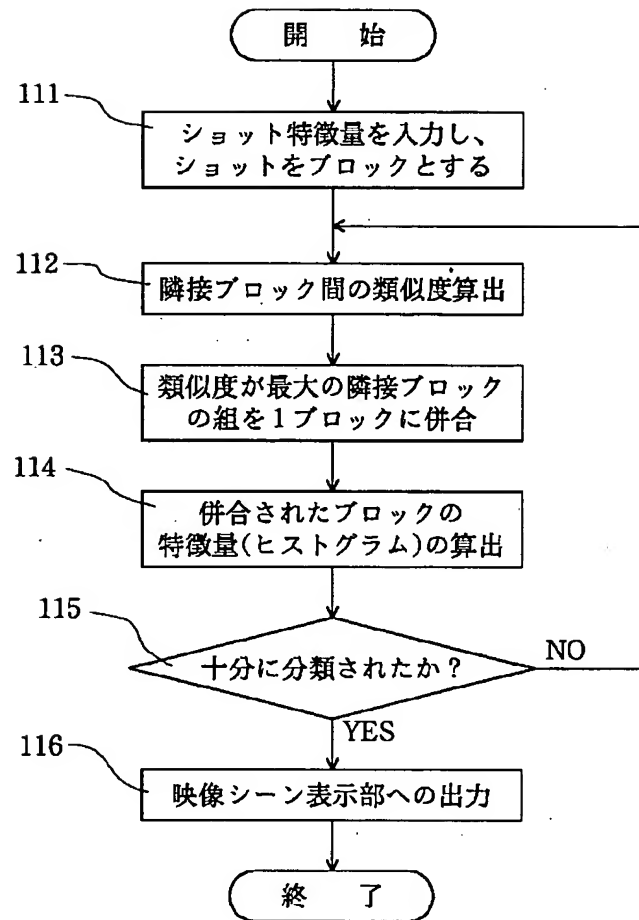
【図3】



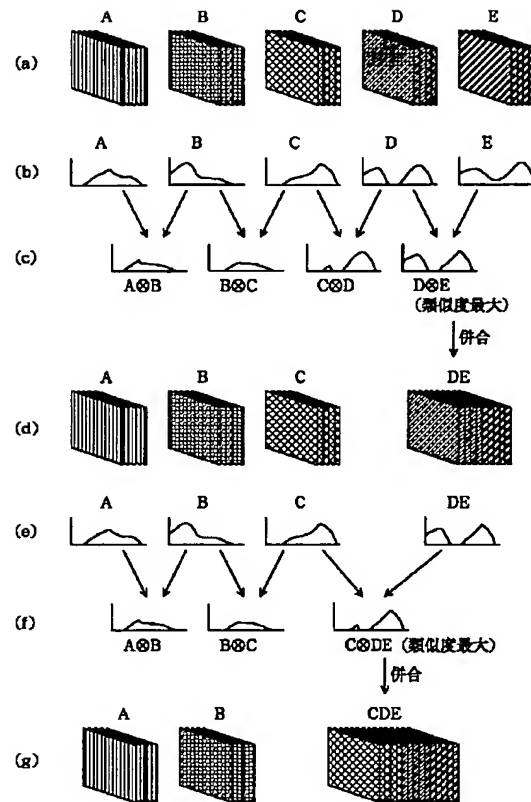
【図5】



【図4】



【図6】



フロントページの続き

(72)発明者 谷口 行信  
 東京都千代田区内幸町一丁目1番6号 日  
 本電信電話株式会社内